

ЛЕКЦІЯ 5. ЗНАХОДЖЕННЯ ОБЛАСТІ ВИЗНАЧЕННЯ МОДЕЛІ

НАВЧАЛЬНІ ПИТАННЯ ТЕМИ ТА ОСНОВНІ ТЕРМІНИ

Область та псевдообласть визначення моделі

Межі умовних інтервалів варіації пояснюючих ознак

1. Область та псевдообласть визначення моделі

Регресійні моделі найчастіше застосовуються для визначення прогнозного значення результуючої ознаки при заданому наборі чинників. Однак ці значення не можна обирати довільно, вони повинні відноситись до певної області, яка визначається значеннями реалізацій ознак. Ця область називається областю визначення моделі. Довести, що задана точка належить цій області є досить складним завданням. В такому випадку не можна обмежитись перевіркою умов

$$a_j \leq x_{ij} \leq b_j, \quad (5.1)$$

де $a_j = \min x_{ij}$, $b_j = \max x_{ij}$,

оскільки така перевірка не є достатньо точною. Наприклад, для двовимірного простору (площини) система обмежень (5.1) задасть прямокутну область (рис 5.1), у той час як область визначення моделі значно менша і являє собою еліпсоїдальну фігуру.

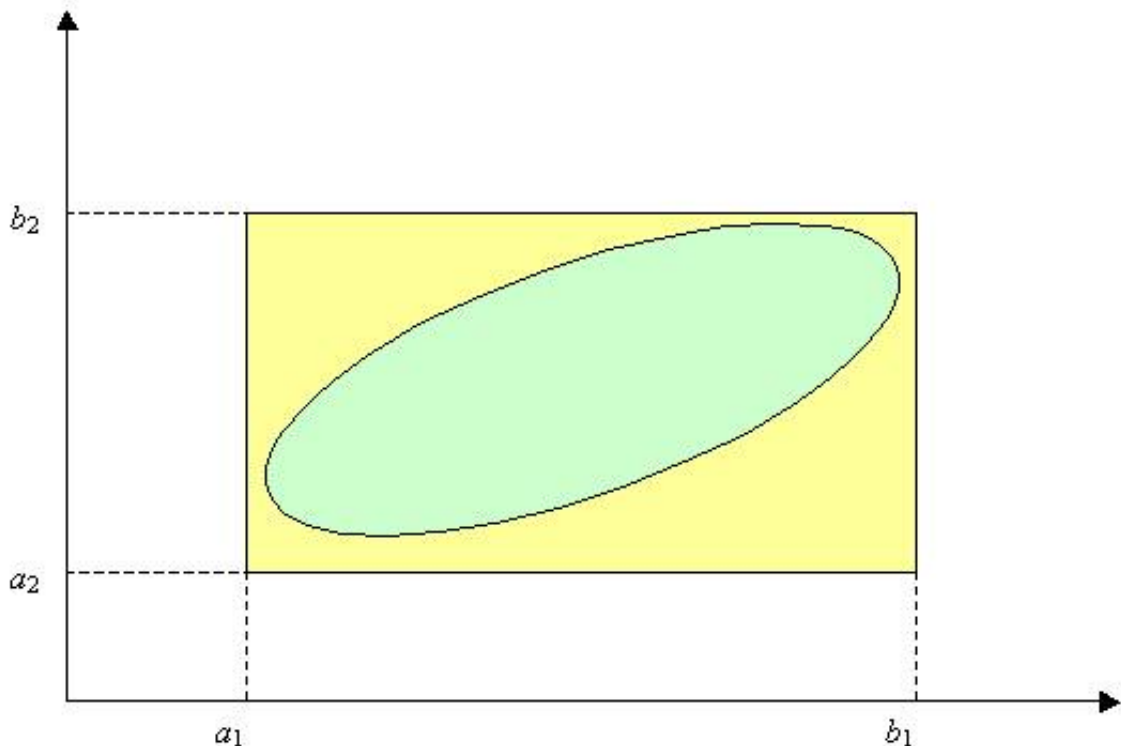


Рисунок 5.1 – Область та псевдообласть визначення моделі

Прямокутний багатомірний паралелепіпед, що одержується внаслідок обмежень (5.1), називається псевдообластю визначення моделі.

Очевидно, що точно визначити область визначення моделі важко і не завжди можливо. В громіздких моделях при великій кількості пояснюючих

ознак обчислення прогнозних значень пов'язано з помилками, оскільки вихідні дані для прогнозу можуть обиратись із псевдообласті визначення моделі.

Для знаходження області визначення моделі необхідно знайти межі реалізацій пояснюючих ознак. Для цієї мети скористаємось поняттям полюса та осі сукупності.

Як відомо, вісь сукупності являє собою пряму, що проходить через верхній $P_B(x_1^*, x_2^*, \dots, x_n^*)$ та нижній полюси $P_H(x_1^*, x_2^*, \dots, x_n^*)$. Тоді параметричне рівняння осі сукупності має такий вигляд:

$$x_i = x_{i*} + a_i \times t, \quad (5.2)$$

де $a_i \times = x_{i*}^* - x_{i*}$.

Для знаходження області визначення моделі потрібно знати проекції точок P_i на вісь сукупності, причому проекції повинні бути перпендикулярними одній з осей координат, які утворюють пояснюючі ознаки. Ознака, що визначає цю вісь, обирається на основі змістовного аналізу. Ця ознака буде характеризуватись тим, вона приймає довільні значення на всьому інтервалі варіації. Тому в ролі такої ознаки обирається та, для якої з економічної точки зору це найбільш природно. Цю ознаку називають вільною ознакою. Значення решти ознак одержують значення в залежності від заданого значення вільної ознаки.

Якщо змістовний аналіз не дозволяє однозначно обрати вільну ознаку, то використовують методи статистичного аналізу. В такому випадку вільною

обирають ознаку з найбільшим значенням коефіцієнта варіації $v = \frac{\sigma}{\bar{x}}$.

Позначимо вільну ознаку через X_1 . Наступним кроком після її визначення є побудова прямокутної проекції осі сукупності з простору R^n в простір R^2 , який задається системами координат OX_1X_k . Ці проекції виражаються рівняннями

$$\begin{aligned} z_1 &= x_{1*} + (x_1^* - x_{1*})t \\ z_k &= x_{k*} + (x_k^* - x_{k*})t. \end{aligned} \quad (5.3)$$

Далі через точки $P_s = (x_{s1}, x_{sk})$, які являють собою проекції точок $P_s = (x_{s1}, x_{s2}, \dots, x_{sn})$ на відповідні площини OX_1X_k проводяться прямі, перпендикулярні OX_1 . Рівняння цих прямих мають вигляд $z_1 = x_1$.

Перетин цих прямих з прямими, що задаються рівняннями (5.3) дають точки, які і є шуканими проекціями емпіричних точок на вісь сукупності. Тоді виконується умова $z_1 = x_{s1}$, яка дозволяє знайти параметр t з першого рівняння системи (5.3):

$$t = \frac{x_{s1} - x_{1*}}{x_1^* - x_{1*}}. \quad (5.4)$$

Підставивши значення t в друге рівняння системи (5.3), одержимо другу координату точки W_s :

$$w_{sk} = x_{k*} + (x_k^* - x_{k*}) \frac{x_{s1} - x_{1*}}{x_1^* - x_{1*}}. \quad (5.5)$$

Таким чином одержують проекції W_s точок P_s на вісь сукупності. Позначивши $w_{s1} = x_{s1}$, запишемо координати проекцій цих точок: $W_s = (w_{s1}, w_{sk})$.

Викладена процедура знаходження проекцій проілюстрована на рис. 5.2. Червоним кольором відмічені точки-полюси, яких насправді у вихідній сукупності даних немає.

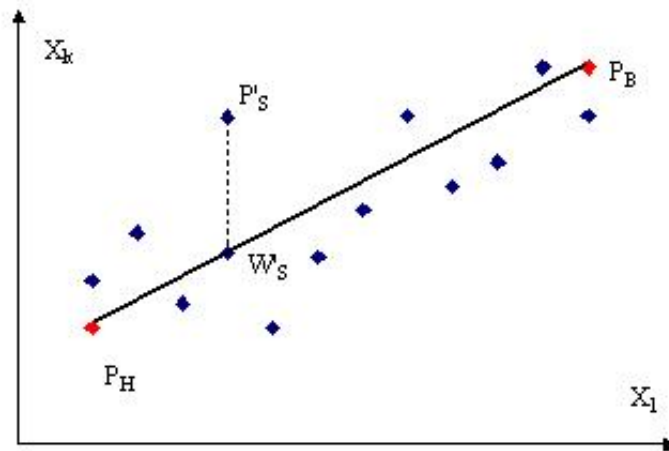


Рисунок 5.2 – Побудова проекцій точок на проекцію осі сукупності

В пропонуваній методиці проекції будуються двічі. Перший раз – коли будуються проекції вихідних точок P_s і осі сукупності на площини OX_1X_k . Другий раз – коли знаходяться проекції точок P_s на проекції осі сукупності; ці дії вже проводяться у двовимірному просторі (що і зображено на малюнку).

1. Межі умовних інтервалів варіації пояснюючих ознак

Знайдені точки P_s та W_s дозволяють обчислити різниці міжкоординатами пояснюючих ознак моделі і, відповідно, визначити розмах варіаційних ознак при довільних значеннях вільної ознаки. Різниці d_{sk} знаходяться за формулою

$$d_{sk} = |x_{sk} - w_{sk}|. \quad (5.6)$$

Ці різниці характеризують ступінь розсіювання точок P_s навколо осі сукупності, причому вимірювання величин різниць здійснюється паралельно осі X_1 . Якщо знайти максимальні значення таких різниць

$$d_k = \max_s d_{sk}, \quad (5.7)$$

то інтервали, довжина яких становить $2d_k$, будуть містити всі точки-спостереження P_s . Використовуючи одержані величини, можна перевіряти приналежність довільної точки до області визначення моделі. Процедура перевірки точки $P_a = (x_{a1}, x_{a2}, \dots, x_{an})$ укладається у виконання наступних

кроків:

1) Знайти проекцію точки $P_a = (x_{a1}, 0, \dots, 0)$ на вісь сукупності.

Одержимо точку $W_a = (w_{a1}, w_{a2}, \dots, w_{an})$. Перша координата знаходиться з

рівності $w_{a1} = x_{a1}$, інші – за формулою (5.5): $w_{ak} = x_{k*} + (x_k^* - x_{k*}) \frac{x_{a1} - x_{1*}}{x_1^* - x_{1*}}, k=2, 3, \dots, n$.

При цьому повинна виконуватись умова $\min x_{s1} \leq x_{a1} \leq \max x_{s1}$.

2) Опираючись на вираз (5.7), знайти допустимі межі умовних інтервалів варіації пояснюючих ознак:

$$\begin{aligned} c_k' &= w_{ak} - d_k \\ c_k'' &= w_{ak} + d_k \\ c_k' &\leq x_{ak} \leq c_k'' \end{aligned} \quad (5.8)$$

Крім того, повинна виконуватись умова

$$c_k' \geq \min_j x_{jk}, \quad c_k'' \leq \max_j x_{jk}. \quad (5.9)$$

Очевидно, що в залежності від вибору вільної ознаки кожен раз буде побудована нова область визначення моделі. Всі одержані таким чином розв'язки будуть рівнозначними з тієї точки зору, що межі інтервалів варіації ознак в різних варіантах не суперечать один одному. Це означає, що коли деяка точка P не належить області визначення моделі в одному якомусь варіанті, то вона не буде їй належати і в інших варіантах. І навпаки, якщо така точка входить в область визначення моделі в одному варіанті, то вона увійде до неї і в інших варіантах також.

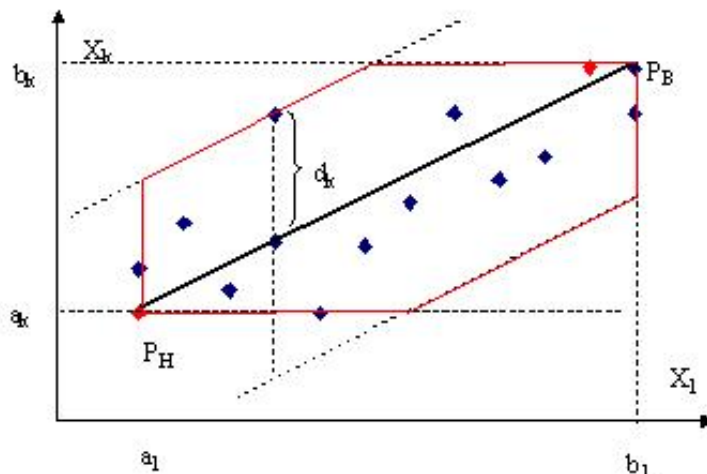


Рисунок 5.3 – Побудова області визначення моделі

На рис 5.3 наведений приклад побудови фрагмента області визначення моделі в одній з площин. Як випливає з рисунка, вона дійсно має еліпсоїдальну форму. Помітимо, що якби в ролі вільної ознаки був би обраний чинник X_k , то ми би одержали таку саму фігуру.

В загальному випадку область визначення моделі геометрично буде однаковою для кожного варіанта її побудови, і умовні інтервали варіації ознак будуть несуперечливими.

Приклад.

Завдання 5.1.

Для аналізу залежності продуктивності праці (Y) від наступних чинників:

X_1 – вартості сировини та матеріалів (млн. грн.);

X_2 – заробітної плати (тис. грн.);

X_3 – основних промислово-виробничих фондів (млн. грн.);

X_4 – відрахувань на соціальне страхування (тис. грн.)

побудована економетрична модель

$$Y = 15,031 + 0,883 X_1 + 2,876 X_2 - 0,006 X_3 - 0,076 X_4$$

Дані для побудови моделі наведені у табл. 5.1. Потрібно визначити, чи можна використати цю модель для обчислення прогностичних значень продуктивності праці при заданих значеннях чинників, які наведені у табл. 5.2.

Таблиця 5.1

№	X_1	X_2	X_3	X_4	Y
1	0,865	0,651	2,627	54	10,6
2	9,571	1,287	9,105	105	19,7
3	1,334	1,046	3,045	85	17,7
4	6,944	0,944	2,554	79	17,5
5	4,662	1,26	6,417	105	14,4
6	2,1	1,212	4,845	101	9,4
7	1,215	0,254	0,923	19	11,9
8	5,191	1,795	9,602	150	13,9
9	4,965	2,851	12,542	240	8,9
10	2,067	1,156	6,718	96	14,5

Таблиця 5.2

№	X_1	X_2	X_3	X_4
1	15,280	6,175	14,009	405
2	1,020	0,766	4,257	90
3	13,324	1,099	6,021	145

Розв'язок.

Знайдемо координати верхнього та нижнього полюса. Запишемо їх до табл. 5.3

Таблиця 5.3

P_H	0,865	0,254	0,923	54	8,9
P_B	9,571	2,851	12,542	240	19,7

Оберемо в ролі вільної ознаки чинник X_1 . Далі обчислимо координати проекцій точок на проекції осі сукупності для кожної площини. Значення проекцій наведені в табл. 5.4.

Таблиця 5.4

W_1	W_2	W_3	W_4
0,865	0,254	0,923	54,000
9,571	2,851	12,542	240,000
1,334	0,394	1,549	64,020
6,944	2,067	9,036	183,875
4,662	1,387	5,990	135,121
2,100	0,622	2,571	80,385
1,215	0,358	1,390	61,478
5,191	1,544	6,696	146,423
4,965	1,477	6,395	141,595
2,067	0,613	2,527	79,680

За даними таблиць 5.1 та 5.4 обчислимо значення відхилень $d_{s,k} = |x_{s,k} - w_{s,k}|$. Результати занесемо до табл. 5.5. Визначимо максимальні значення цих різниць d_k і результат запишемо також до табл. 5.5.

Таблиця 5.5

D			
0	0,397	1,704	0
0	1,564	3,437	135
0	0,6521	1,49607	20,98
0	1,1234	6,48201	104,875
0	0,1266	0,42654	30,1213
0	0,5896	2,27377	20,6147
0	0,1044	0,46711	7,5224
0	0,2506	2,90553	3,57684
0	1,374	6,14715	98,4052
0	0,5434	4,19081	16,3198
d_k	1,564	6,48201	135

Проведемо аналіз значень чинників, для яких потрібно обчислити прогноз. Оскільки для значень чинників з першого набору не виконується умова $\min x_{s1} \leq x_{s1} \leq \max x_{s1}$. Для другого та третього наборів обчислимо значення координат точок W_k (табл. 5.6) і різниць d_k (табл. 5.7).

Аналіз таблиці 5.7 показує, що для значень другого набору виконуються всі умови приналежності точки до області визначення моделі, тому для неї можна розрахувати прогнозне значення. В даному випадку воно дорівнює $Y_{\text{ГР}} = 11,22$. Значення чинників третього набору виходять за межі області визначення,

тому розраховувати для них прогноз некоректно.

Таблиця 5.6

№	W_1	W_2	W_3	W_4
2	1,020	0,300	1,130	57,312
3	9,324	2,777	12,212	234,723

Таблиця 5.7

№	D			
2	0	0,4658	3,12714	32,6885
3	0	1,6783	6,19135	89,7229

2. Питання для самоперевірки.

1. Поясніть, що розуміють під областю та псевдообластю визначення моделі. В чому між ними відмінності?
2. Для чого використовують вісь сукупності при побудові області визначення моделі?
3. Що розуміють під вільної ознакою при побудові області визначення моделі? Як вона обирається?
4. Що являють собою межі умовних інтервалів варіації ознак? Як вони будуються?